

University of Groningen

## Revealing Causal Heterogeneity Using Time Series Analysis of Ambulatory Assessments

Rosmalen, J.G.M.; Wenting, A.M.G.; Roest, A.M.; de Jonge, P.; Bos, Elisabeth

*Published in:*  
Psychosomatic Medicine

*DOI:*  
[10.1097/PSY.0b013e3182545d47](https://doi.org/10.1097/PSY.0b013e3182545d47)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2012

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Rosmalen, J. G. M., Wenting, A. M. G., Roest, A. M., de Jonge, P., & Bos, E. (2012). Revealing Causal Heterogeneity Using Time Series Analysis of Ambulatory Assessments: Application to the Association Between Depression and Physical Activity After Myocardial Infarction. *Psychosomatic Medicine*, 74(4), 377-386. <https://doi.org/10.1097/PSY.0b013e3182545d47>

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# Revealing Causal Heterogeneity Using Time Series Analysis of Ambulatory Assessments: Application to the Association Between Depression and Physical Activity After Myocardial Infarction

JUDITH G.M. ROSMALEN, PhD, ANGELA M.G. WENTING, MSc, ANNELIEKE M. ROEST, MSc, PETER DE JONGE, PhD,  
AND ELISABETH H. BOS, PhD

**Objective:** Studies in psychosomatic medicine are characterized by analyses that typically compare groups. This nomothetic approach leads to conclusions that apply to the average group member but not necessarily to individual patients. Idiographic studies start at the individual patient and are suitable to study associations that differ between time points or between individuals. We illustrate the advantages of the idiographic approach in analyzing ambulatory assessments, taking the association between depression and physical activity after myocardial infarction as an example. **Methods:** Five middle-aged men who had myocardial infarction with mild to moderate symptoms of depression were included in this study. Four of these participants monitored their physical activity and depressive symptoms during a period of 2 to 3 months using a daily self-registration form. The time series of each individual participant were investigated using vector autoregressive modeling, which enables the analysis of temporal dynamics between physical activity and depression. **Results:** We found causal heterogeneity in the association between depression and physical activity. Participants differed in the predominant direction of effect, which was either from physical activity to depression ( $n = 1$ , 85 observations, unstandardized effect size =  $-0.183$ ,  $p = .03$ ) or from depression to physical activity ( $n = 2$ , 65 and 59 observations, unstandardized effect sizes =  $-0.038$  and  $-0.381$ ,  $p < .001$  and  $p = .04$ ). Also, the persistency of effects differed among individuals. **Conclusions:** Vector autoregressive models are suitable in revealing causal heterogeneity and can be easily used to analyze ambulatory assessments. We suggest that these models might bridge the gap between science and clinical practice by translating epidemiological results to individual patients. **Key words:** time series analysis, ambulatory assessment, depression, physical activity, vector autoregressive models.

PEP = Psycho-Educational Prevention Module; BDI = Beck Depression Inventory; PCI = percutaneous coronary intervention; CABG = coronary artery bypass graft; LVEF = left ventricular ejection fraction; BMI = body mass index; VAR = vector autoregressive modeling.

## INTRODUCTION

The heart of psychosomatic medicine lies in the interaction between psychosocial and biomedical factors in the etiology of and coping with disease. Typical studies in this field compare groups of patients and healthy controls on psychosocial factors (case-control studies), identify risk factors by following populations for the development of their health (cohort studies), or allocate patients to treatment or placebo to study the effect of a treatment program (intervention studies). What these studies have in common is their nomothetic approach. This approach involves studying groups of individuals with a sample size as large as possible, aggregating data from these individuals and presenting results in group averages, to study an association between variables at the population level (1). Provided the sample is representative and large and the association is replicated, the generalizability of the results to the population is high.

However, the generalizability of these results to *specific individuals* is often low. One may wonder what the use is for an individual patient, to know that psychosocial stress accounts for approximately 30% of the attributable risk of acute myocardial infarction (MI) (2) or that antidepressant medication ex-

plains 1% to 4% of the reduction in depressive symptoms on top of placebo (3). Even major effects that are present on the group level do not necessarily apply to an individual patient. Reversibly, small or absent effects at the population level are usually interpreted as clinically irrelevant, but these may be of crucial importance in particular patients. The finding that only a very small percentage of variance in depressive symptoms might be attributed to inflammation (4) does not preclude the possibility that, for individual patients, depression may be fully induced by inflammatory markers.

In the nomothetic approach, problems of external validity of study results are well recognized; the generalizability of results to the population usually composes a significant part of the discussion of articles. This is in sharp contrast to the generalizability of results obtained at the group level to individual patients. Often, there is unjustified jumping from the level of the group to the level of the individual patient. Such generalization of group-level results to the individual level is not automatically valid. First, a between-subjects correlation is not necessarily the same as a within-subject correlation (5,6). An association found at the population level may prove nonexistent or even reversed at the individual level (7). Second, many variables studied in the field of psychosomatic medicine show large intraindividual variability. Factors such as hypothalamic-pituitary-adrenal (HPA) axis and autonomic nervous system activity, life-style factors such as physical activity, and psychological factors like depression, are not static but dynamic. They typically show large fluctuations over time within individuals. This represents a problem for the nomothetic research approach because the number of measurement points in conventional group studies is usually very limited, and the interval between measurements is large. Third, there is a large interindividual variability in the presentation and dynamics of these variables. Studies on hypothalamic-pituitary-adrenal axis show that

From the University of Groningen (J.G.M.R., P.d.J., E.H.B.), Groningen; and Tilburg University (A.M.G.W., A.M.R.), Tilburg, the Netherlands.

Address correspondence and reprint requests to Peter de Jonge, PhD, Department of Psychiatry, University of Groningen, Groningen, the Netherlands. E-mail: peter.de.jonge@umcg.nl

Received for publication April 28, 2011; revision received November 2, 2011.

DOI: 10.1097/PSY.0b013e3182545d47

cortisol levels are peeking in the early morning, but this is not necessarily true for all individual patients that this averaged curve is composed of (8). Differences between individuals in life-style behaviors are obviously large as well. And also with regard to psychological processes, between-subjects heterogeneity is the norm rather than the exception. Because psychosomatic medicine is the branch of science that focuses on the patient instead of on the disease, we are all aware of individual differences. What we do not realize is that the nomothetic approach we are using deals with variability between individuals as if it were error.

We present an alternative: an idiographic approach, which is based on analyses within individuals (1,6). These analyses are performed on multiple repeated measurements (time series) of the variables of interest within a single individual. Instead of explaining variance in the population, the aim of this approach is to explain variance within single individuals without assuming that individuals are interchangeable. Idiographic studies using time series analysis have several advantages. First, differences between individuals are easier to detect because results are analyzed at the individual level. Between-subjects heterogeneity is easily obscured when results are averaged over a large number of subjects. Second, the temporal dynamics of the association of interest are more adequately investigated thanks to the multitude of repeated measurements separated by small time lags. This makes these studies very suited for investigating time-lagged influences and the temporal ordering of effects. It is not necessary to a priori decide which variable is the predictor and which variable is the outcome, and bidirectional effects and feedback cycles can be modeled as well. These options are very difficult or in fact impossible to analyze appropriately in most nomothetic designs. Third, the results have high ecological validity because the data are typically collected in daily life using real-time ambulatory assessments. Fourth, the data have more clinical relevance to individuals than data collected using nomothetic designs. They apply to concrete individuals in specific contexts and enable a patient-tailored advice.

In this article, we will apply an idiographic analysis to the association between depression and physical activity after MI. Depression and lack of physical activity after MI are two interrelated risk factors for poor cardiovascular prognosis. Previous studies have suggested a negative association between depression and physical activity, but the directionality and temporal dynamics of this relationship remain unclear. A recent observational study studied depression and physical activity in a group of elderly adults and found that physical activity mediated the relationship between depression and cardiovascular events and mortality (9). At the same time, physical activity is tested as a treatment of depression (10). We will analyze the depression–physical activity association in individuals using vector autoregression, a novel time series analysis approach (11). We will show that the relevant question is not whether there is an association between depression and physical activity in general, but what this association looks like in individuals. This latter question allows the elucidation of the sequence of events leading to the development of symptoms in individuals.

## METHODS

### Study Design

We adopted an idiographic approach, investigating the microstructure of the dynamic relationship between depressive symptoms and physical activity in five individuals after an MI. Participants included were post-MI patients recruited from screening for a Psycho-Educational Prevention Module (PEP) at Máxima Medical Center, Eindhoven-Veldhoven, the Netherlands. This module focuses on regaining emotional stability and dealing with cardiac disease, as part of a cardiac rehabilitation program. To test their eligibility for the current study, the patients were screened for depressive symptoms using the Beck Depression Inventory (BDI) (12). This self-report questionnaire measures depressive symptoms in a reliable and valid manner (13). The questionnaire assesses cognitive as well as somatic depressive symptoms during the past week, such as hopelessness, guilt, fatigue, and weight changes. The BDI total score is the sum of all 21 items, each being scored on a scale ranging from 0 to 3. The cutoff score for participation in this study was a BDI score of 10 or higher, which is indicative of at least mild to moderate symptoms of depression. Exclusion criteria were significant cognitive impairments, life-threatening diseases, and severe problems with physical activity. Written informed consent was obtained from all participants. The study was approved by the Medical Ethical Committee for mental health institutions in the Netherlands. Data collection took place in the first semester of 2010. Participants were asked to monitor their daily levels of physical activity and depressive symptoms during a 3-month period, starting while participating in the PEP program. To encourage compliance, we told the participants that we would provide a personal report of the results by using the data they would record themselves (and we did) and participants felt this to be quite compelling. Moreover, we had weekly contacts with the participants in which we reviewed the data recording and offered the participants a small financial incentive (25 euros after completing the series).

### Measures

Physical activity and depressive symptoms of the participants were monitored using a daily self-registration form. Depressive symptoms were assessed by means of the Depression module of the Patient Health Questionnaire (14), which we adapted for daily use. The Patient Health Questionnaire includes nine items assessing depressive symptoms directly based on the nine DSM-IV criteria for major depressive disorder. The items are rated on a 4-point scale from 0 (not at all) to 3 (very much). We used the sum score (0–27) as a measure of depression severity. Sum scores of 10 or higher are considered as an indication of clinically relevant depressive symptoms (14).

The degree of physical activity was assessed by seven items measuring commuting activities, work activities, household activities, sports, and other leisure-time activities. Examples were mentioned to illustrate which type of activities were applicable. Participants registered the amount of time in minutes they had spent on these physical activities. We used the total daily amount of time of physical activity, including both light to moderate and intensive activities, for our analyses.

### Demographic and Clinical Characteristics

Clinical variables were obtained from the patients' medical records. These included cardiac history (MI, percutaneous coronary intervention (PCI), coronary artery bypass graft (CABG) before the index MI), left ventricular ejection fraction (LVEF), medical comorbidities, and medication use. Body mass index (BMI), smoking status, and prior psychological counseling were obtained using self-report. Demographic variables included age, sex, partner status, and educational level were also obtained using self-report.

### Statistical Analysis

The time series of each individual participant were investigated using a technique for the analysis of multiple time series called vector autoregressive (VAR) modeling (11,15). The VAR technique was originally developed by Sims (16) for research in econometrics. It has now also been used in fields like meteorology, sociology, political sciences, and neuroimaging. Vector autoregressive modeling is especially suitable for investigating the temporal dynamics between two or more time series. An attractive feature of this technique is its ability to investigate causal relationships between variables. By separating the dynamic

## TIME SERIES ANALYSIS OF AMBULATORY ASSESSMENTS

part of the model (the longitudinal part, i.e., the relationships between the time-lagged values of the variables) from the simultaneous part (the cross-sectional part, i.e., the relationships between the contemporaneous values), the model allows to make inferences about the temporal order of the effects and thus about causality (11). A further advantage is that VAR modeling allows for bidirectionality and feedback effects. This is important because relationships between variables studied in psychosomatic medicine are often not unidirectional but may mutually influence each other.

A VAR model is a multivariate autoregressive model that consists of a set of regression equations for a system of two or more variables (11). All variables in the system are treated as endogenous, which means that they can be both determinant and outcome. In the present study, a two-variable VAR modeling was used. This VAR model consisted of a system of two endogenous variables, namely, depression and activity, and included the following regression equations:

$$Activity_t = \alpha_0 + \sum_{i=1}^p \alpha_i Activity_{t-i} + \sum_{i=1}^p \beta_i Depression_{t-i} + \varepsilon_{1t}$$

$$Depression_t = \beta_0 + \sum_{i=1}^p \gamma_i Activity_{t-i} + \sum_{i=1}^p \delta_i Depression_{t-i} + \varepsilon_{2t}$$

where  $\alpha_i$ ,  $\beta_i$ ,  $\gamma_i$ , and  $\delta_i$  are the coefficients to be estimated,  $p$  is the number of lags considered in the system, and the  $\varepsilon_t$ 's are the stochastic error terms. Each of the two endogenous variables is regressed on its own  $p$  lagged values and the  $p$  lagged values of the other variable. The error terms are called innovations or shocks in the language of VAR modeling. These errors should be serially uncorrelated but can be contemporaneously correlated. To account for the effects of potential confounding variables, control variables can be added to the VAR model. Such control variables are exogenous to the system (which means that they may influence the system but they cannot be influenced by the system). The best-suited number of lags that is needed in the model (i.e., the "VAR order") can be determined using lag length selection criteria such as the likelihood ratio test (LR test), final prediction error (FPE), Akaike Information Criterion (AIC), Hannan-Quinn Information Criterion (HQIC), and Bayesian Information Criterion (BIC). The optimum lag length is the one that minimizes goodness-of-fit statistics (15).

After estimation of the VAR modeling, the coefficients of parameters not contributing to the model can be constrained (set to 0) (15). The VAR model is

reestimated after placing each constraint. Fit statistics can be used to compare the fit of successive models. We used the BIC for this purpose. If this criterion did not indicate a worsening of model fit, the constraint was retained. Parameters with the lowest  $t$  values were constrained first.

A number of diagnostic tests were performed to check whether the final models were correctly specified (15). To establish the stability of the models, we tested the eigenvalue stability condition. The white noise assumption (no residual autocorrelation) was tested by means of Portmanteau tests and inspection of the autocorrelation functions. Homoscedasticity (variance stationarity) was tested by means of Portmanteau tests on the squares of the residuals. The assumption of normally distributed residuals was tested using the Jarque-Bera test, the skewness test, and the kurtosis test. If one of these tests indicated a violation of the model assumptions, the model was adjusted, reestimated, and reevaluated, in an iterative model building process, until all assumptions were met.

In VAR modeling, the regression coefficients cannot be interpreted individually because it is the behavior of the system and all its coefficients that describe the dynamics of the variables (11). Therefore, VAR modeling is usually accompanied by the techniques of Granger causality testing, impulse response function (IRF) analysis, and forecast error variance decomposition, which give an indication of the system's dynamic behavior.

Granger causality is a test for the directionality of the influence between two time series. The essential idea behind Granger causality is that a cause cannot come after an effect (17). This temporal ordering of events can be used to empirically distinguish between leading and lagging variables. A variable  $Y$  is said to "Granger cause"  $Z$  if past values of  $Y$  improve the prediction of  $Z$ , and more so than past values of  $Z$  alone can do (15).

Impulse response functions allow tracing out the dynamic impacts of changes in each of the endogenous variables over time. They do so by visualizing the influence of an isolated shock in one of the variables to the other variable(s), showing how this shock is fed through the system. The IRF is based on the VAR model and thus takes into account the time-lagged relationships between the variables only. Orthogonalized IRFs (OIRFs) are variants of IRFs that take into account the contemporaneous correlations as well (11). These correlations represent the simultaneous relationships between depression and activity, that is, between the scores on these variables within the same measurement day. Here, a problem arises because the order of the effects within measurement days is

TABLE 1. Demographic and Clinical Characteristics of the Study Participants

Characteristic	Participant 1	Participant 2	Participant 4	Participant 5
Age	55	59	51	59
Sex	Male	Male	Male	Male
Partner status	Married	Married	Married	Married
Educational level	Lower vocational	Vocational	Lower vocational	Vocational
Cardiac history	PCI	PCI	PCI	PCI
LVEF	55	72	36	25
Medical comorbidity	None	None	None	Epilepsy
BMI, kg/m <sup>2</sup>	30	25	35	30
Smoking status	Smoker	Nonsmoker	Nonsmoker	Nonsmoker
Medication use	Acetylsalicylic acid Plavix Metoprolol Statine	Acetylsalicylic acid Metoprolol Clopidogrel Amlodipine Isosorbidedi-nitrat Isosorbidemono-nitrat	Acenocoumarol Metoprolol Perindopril Isosorbidedi-nitrat	Acenocoumarol Carvedilol Candesartan Furosemide Statine
BDI at intake	24	17	24	14
Length of time series, d	85	64	65	59
Depressive symptoms, M (SD)	11.8 (2.1)	5.1 (1.9)	13.9 (3.0)	6.6 (2.1)
Physical activity, M (SD), min	431.3 (136.6)	93.3 (80.5)	52.3 (47.9)	126.9 (76.0)

PCI = percutaneous coronary intervention; LVEF = left ventricular ejection fraction; BMI = body mass index; BDI = Beck Depression Inventory; M = mean; SD = standard deviation.

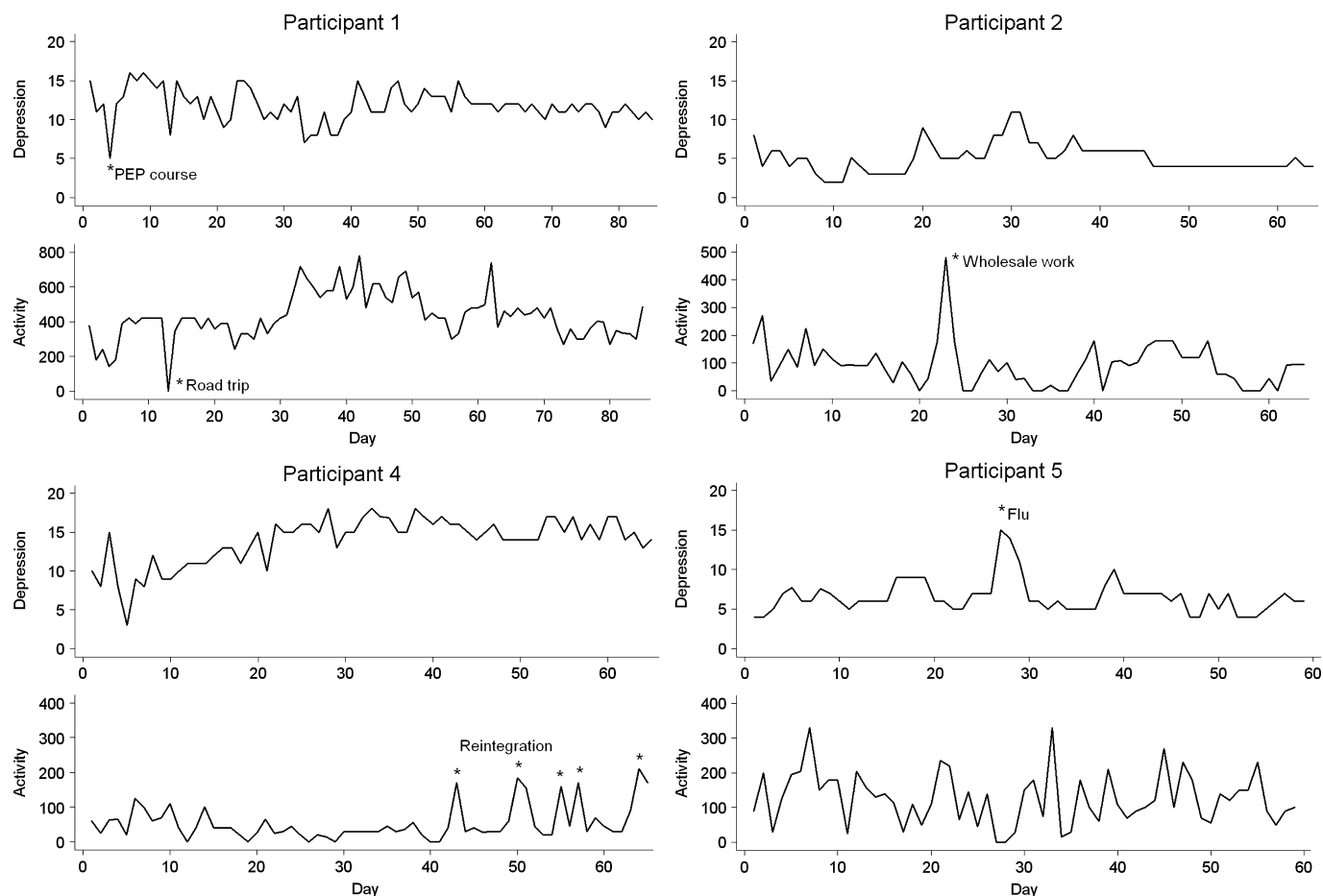


Figure 1. Daily ratings of depressive symptoms (Patient Health Questionnaire 9; range = 0–27) and physical activity (minutes) for Participants 1, 2, 4, and 5. Asterisks refer to outliers due to unusual events recorded in the diary that challenged model assumptions (see text).

unknown. In other words, it is unclear if changes in depression precede or follow same-day changes in activity. Orthogonalized IRFs assume that a specific ordering is chosen for the direction of this contemporaneous relationship. If no theory is available guiding this choice, the results of alternative orderings can be presented. The critical point in VAR modeling is that the decision about this ordering can be made explicit and can be evaluated *after* accounting for the dynamics in the data (11).

Forecast error variance decomposition, or “variance decomposition” (VD), is a final tool for interpreting the results of a VAR model. This technique is useful for estimating the amount of variance in each variable that can be explained by the other variable(s) during a specific period. It gives an impression of the relative influence of the variables in the system (11).

A two-tailed  $\alpha$  level of 0.05 was used to determine statistical significance. Although Sims and Zha (18) recommend the use of 68% error bands for IRFs, we applied the more conservative 95% levels also for this purpose. Analyses were performed in STATA 11 using the suite of VAR commands (19). Maximum likelihood estimation with a degrees-of-freedom adjustment advocated for small samples (15) was used in estimating the VAR coefficients.

## RESULTS

### Descriptives

One of the participants (Participant 3) dropped out the study after 2 weeks, because he had a very busy period at his work and did not manage to keep the diary anymore. We did not analyze the data of this participant. Table 1 summarizes the demographic and clinical characteristics of the other four patients who had MI. All participants were married men with a history

of PCI. They were aged between 51 and 59 years and had a BMI that indicates overweight to obesity. Participant 1 is a smoker and a forest ranger who guides school classes in his spare time. Participant 2 is a former owner in the catering industry, who spends his days in social activities such as playing billiards with friends. Participant 4 is reintegrating into his work as a short-distance truck driver. Participant 5 is a hardworking agricultural owner.

Figure 1 shows the time series of the participants’ daily levels of depressive symptoms and physical activity. The participants monitored their depression and activity levels during 59 to 85 days. Participants 1, 4, and 5 filled out the diary on each day of their series, although some items were missing on some days. The series of Participant 2 contained 3 days on which all item scores were missing. None of the series contained more than 5% missing values. We imputed missing item scores by means of maximum likelihood estimation before computing the depression and activity sum scores.<sup>1</sup>

<sup>1</sup>It is possible to run VAR models with missing values. However, because lagged values cannot be calculated for these missing observations, this may result in a considerable reduction of the sample size if the VAR order is large. Given the fact that our sample sizes were rather small already, we chose to impute the missing values.

## TIME SERIES ANALYSIS OF AMBULATORY ASSESSMENTS

The depression scores of Participant 1 varied from 5 to 16, indicating mild to moderate symptoms. Participant 4 also showed depression scores in the mild to moderate range (3–18). The depression scores of Participants 2 were below the cutoff (10) most of the time (range = 2–11). The same was true for Participant 5 (range = 0–10), except for a short period halfway the series in which he had flu. The average daily levels of depression are presented in the bottom part of Table 1.

The amount of physical activity showed a rather volatile pattern within participants and also varied considerably across participants. Participant 1 showed the highest levels of activity (range = 0–13 hours), which had to do with his work as a forester. Also on nonworking days, this man often went into the forests to show around school classes, as a pastime. Participant 2 showed moderate levels of physical activity except for one day (Day 23), on which he recorded 480 minutes of work-related activity. On this day, he had worked as a substitute in the liquor wholesale where he used to work before his retirement, lifting kegs. The activity series of Participant 4 showed some large peaks in the last phase of the series. During this period, this man was reintegrating into his work as a truck driver for a few days a week in the context of the cardiac rehabilitation program. On these days, he recorded unusually high levels of activity because part of the work was loading his truck.

Because the reasons for the exceptional scores in the series of Participants 2, 4, and 5 were clearly exogenous, we added control variables to the VAR models of these participants (coded 1 on the exceptional days and 0 otherwise). This was also necessary to eliminate violation of the model assumptions (nonnormality and heteroscedasticity). Three additional influential data points were detected during the model building process. The model for Participant 1 showed nonnormality in the residuals of depression and activity owing to exceptional low values at Days 4 and 13, respectively. In his diary, he had made a notation that the first meeting of the PEP course was at Day 4, whereas at Day 13, he had had a road trip looking for a new chicken coop. The depression residuals of Participant 4 showed an outlier at Day 5. No special events were recorded in the diary for this day. Inclusion of control variables for these exceptional data points was required to eliminate violation of the model assumptions. In the models of Participants 2 and 5, we used natural logarithms of the depression and activity scores to correct violation of the normality and homoscedasticity assumptions.

### VAR Lag Order Selection

We determined how many time lags were needed in the VAR models using the lag length selection criteria (LR test, FPE, AIC, HQIC, and SBIC). For the model of Participant 1, all criteria suggested an optimum lag length of 2, which implies that the values of the previous 2 days contained relevant information for current values. For Participant 2, all criteria suggested a lag length of 1. In the model of Participant 5, all criteria except the LR suggested a lag length of 1, so we used a 1-lag model. For Participant 4, the LR test, FPE, and AIC suggested a lag length of 7, whereas the HQIC suggested a lag

TABLE 2. Estimates for the Two-Lag VAR Model of Participant 1

Variable	Dependent Variables	
	Depression <sub>t</sub>	Activity <sub>t</sub>
Depression <sub>t-1</sub>	0.397***	0
Depression <sub>t-2</sub>	0.090	0
Activity <sub>t-1</sub>	-0.183*	0.447***
Activity <sub>t-2</sub>	0	0.273***
Control Variables		
First PEP course (Day 4)	-8.114***	0
Road trip (Day 13)	-5.424**	-7.196***
R <sup>2</sup>	0.39	0.57

VAR = vector autoregressive modeling; PEP = Psycho-Educational Prevention Module.

Shown are regression coefficients from the VAR model. Coefficients denoted with 0 are constraint parameters. Control variables are exogenous to the system. Activity is scaled in hours. Number of observations = 83 (85 - 2 that are lost in calculation for the two lags).

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ .

length of 2 and the SBIC suggested a lag length of 1. Using VAR models with two lags or one lag for this participant resulted in models that violated the white noise assumption. The autocorrelation function of the activity series showed a clear peak at Lag 7, which means that measurements spaced by seven lags were related to each other. Apparently, the activity levels of this participant were strongly related to the days of the week. We introduced dummies for the days of the week in the activity equation to control for this cyclic trend. The periodicity seemed to be largely accounted for by the dummies for Tuesday and Friday, which were the only days that this participant had recorded sports. In contrast, on Sundays, he had systematically recorded very low levels of activity. The dummies for Sunday, Tuesday, and Friday were sufficient to remove all residual autocorrelation at Lag 7, turning the residuals into white noise.

### Estimation of the VAR Model

Next, the VAR models were estimated and evaluated using residual diagnostics. Table 2 presents the final VAR model for Participant 1, as an example.<sup>2</sup> The two endogenous variables depression and activity were modeled as a function of their own previous values (Lags 1 and 2, i.e., 1 day and 2 days before), the previous values of the other endogenous variable (Lags 1 and 2), and the control variables for the exceptional Days 4 and 13. As can be seen in the table, both the depression and activity series showed important positive autocorrelation; the first lag of these variables significantly predicted their own current values. In the activity series, the second lag was also

<sup>2</sup>The VAR models of the other participants are available on request from the corresponding author.

predictive of current activity levels. Besides these autoregressive effects, which represent the internal dynamics of the series, a cross-lagged relationship between activity and depression was observed; the first lag of activity was negatively related to current depression scores. Thus, higher levels of activity on any 1 day were followed by lower levels of depression the next day. The variables in the model explained 39% of the variance in the depression series and 57% of the variance in the activity series.

In a similar way, the models of the other 3 participants were estimated. In the model of Participant 2, only autoregressive effects were present; the cross-lagged effects between depression and activity were not significant. In the models of Participants 4 and 5, the depression series showed positive autocorrelation, but the activity series did not. Interestingly, lagged levels of depression were related to current levels of activity in the models of these participants. Thus, higher levels of depression on any 1 day were followed by lower levels of activity the next day. The reverse relationship, from activity to depression, was not significant in these models. The variance explained by the models was considerable, although great individual differences were observed. Explained variance in the depression series ranged from 39% to 67%, and explained variance in the activity series ranged from 16% to 88%.

### Granger Causality

We performed Granger causality tests to investigate whether depression Granger caused activity or whether the reverse was true, or both. Table 3 shows the results of these tests for each participant. The table shows that the direction of causality was not the same for all participants. Past activity levels predicted current depression scores for Participant 1, but the reverse was true for Participants 4 and 5. Past depression scores predicted current activity levels in these latter participants. In the model of Participant 2, no Granger causality was present. Whether the impact of the observed effects was positive or negative cannot be derived from the Granger tests but becomes clear from the sign of the estimates in the VAR models (see the rightmost column of Table 3). All effects were negative. Thus, increases in activity were followed by decreases in depression or vice versa.

### Contemporaneous Correlations

For all participants, the contemporaneous correlation between depression and activity was negative. The size of the correlations, according to the suggestions provided by Cohen (20), was small to medium: Participant 1,  $r = -0.232$ ; Participant 2,  $r = -0.300$ ; Participant 4,  $r = -0.262$ ; Participant 5,  $r = -0.172$ . Thus, on days that participants were less active, they were more depressed; or on days that they were less depressed, they were more active. The causal direction of this relationship (and whether this effect was causal at all) could not be established from these analyses because the contemporaneous values were measured on the same day. More insight into the nature of these within-day effects would require a study with shorter measurement intervals.

### Impulse Response Analysis

We calculated OIRFs to trace out the impact of a change in each of the variables on the other over time. We present the OIRFs for Participant 1, as an example. Figure 2 shows the results for the two different orderings of the contemporaneous correlation during a 10-day horizon. In Order 1, it is assumed that changes in activity precede same-day changes in depression. In Order 2, the order is reversed. The top panels of Figure 2 show the response of depression to a shock in activity. The shock refers to an innovation of 1 SD in (the residuals of) activity (i.e., approximately 1.5 hour for this participant). The figure shows that, in both orderings, the activity shock leads to a significant decrease in depression for the next 3 days. Thereafter, the response slowly tapers off to become 0 after approximately 12 days. In Order 1, the response is immediate and larger than in Order 2, which is due to the extra effect of the contemporaneous correlation between activity and depression in this ordering. In Order 2, the effect of activity on depression is initially 0, which is a natural result of the chosen ordering. The favorable effect of activity on depression becomes visible only after 1 day in this ordering. The prolonged effect in both OIRFs can be explained by the autoregressive effects in the depression and activity series, which are responsible for the persistency in these variables. The feedback loop from activity to depression also contributes to the prolongation of the effect.

TABLE 3. Granger Causality Tests

Participant	Causality Test	$\chi^2$	df	p	VAR Parameter	Estimate
Participant 1	Activity $\rightarrow$ depression	5.02	1	.03	Activity <sub>t-1</sub>	-0.183
	Depression $\rightarrow$ activity	—	—	—		
Participant 2	Activity $\rightarrow$ depression	—	—	—		
	Depression $\rightarrow$ activity	—	—	—		
Participant 4	Activity $\rightarrow$ depression	—	—	—	Depression <sub>t-1</sub>	-0.038
	Depression $\rightarrow$ activity	31.8	2	<.001	Depression <sub>t-2</sub>	-0.041
Participant 5	Activity $\rightarrow$ depression	—	—	—		
	Depression $\rightarrow$ activity	4.15	1	.04	Depression <sub>t-1</sub>	-0.381

Tests denoted with (—) were not performed because the parameters involved were constrained. A significant  $\chi^2$  value implies that the first variable “Granger causes” the second variable. Thus, the first variable is considered to have causal impact on the second variable. Whether the impact is positive or negative can be derived from the sign of the estimates in the VAR models (right column).

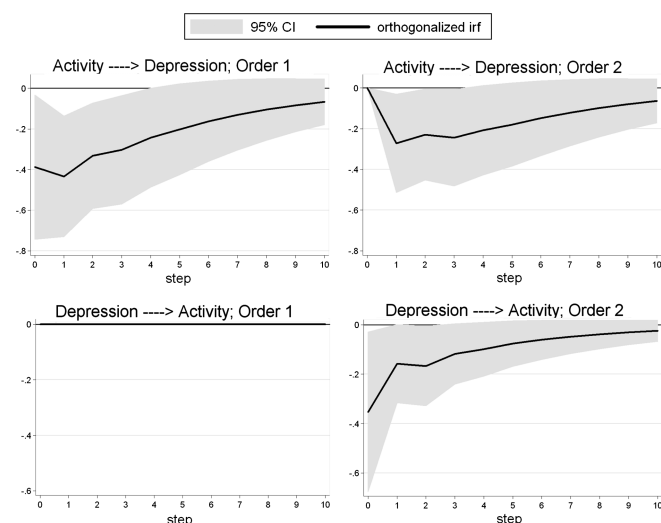


Figure 2. Orthogonalized impulse response functions (IRFs) for Participant 1. Top panels: impulse = activity; bottom panels: impulse = depression. Responses are considered significant if their error bands do not include 0. Responses are plotted during a 10-day horizon. Order 1 = activity – depression; Order 2 = depression – activity; CI = confidence interval.

The bottom panels of Figure 2 shows the OIRFs for the response of activity to a shock in depression. The shock here refers to an innovation of 1 SD in depression (approximately 1.6 units for this participant). In Order 1, the activity response is 0 over the whole horizon, which is a natural result of the chosen ordering for the contemporaneous correlations and the fact that no lagged effects from depression to activity were present in the system of this participant. If Order 2 is assumed, an immediate and significant decrease in activity is observed, which is initially entirely driven by the contemporaneous relationship between depression and activity. This effect has largely disappeared the next day, although some prolongation of the effect is established at later days owing to the autoregressive effects in the variables and the feedback effect from activity to depression.

In summary, the ordering of the variables seems to have little impact on the general conclusion that past activity levels influence current depression scores in this participant. The ordering does have impact on the conclusion about the effect of depression on activity, which would be nonexistent in Order 1 and short term but existent in Order 2.

### Cumulative IRFs

The total impact over time of a shock in one of the variables on the other is calculated by computing the cumulative IRF. The results for Participant 1 are summarized in Table 4. The leftmost column of Table 4 shows that a 1-SD shock to activity leads to a total decrease in depression of 2.45 units during a period of 10 days in Order 1. The total decrease is 1.65 units if Order 2 is assumed (second column). The third and fourth columns show the cumulative IRF for the response of activity to a shock in depression. In Order 1, this response is 0. If Order 2 is assumed, a total decrease of 1.17 points in activity (i.e., approximately 70 minutes) is observed during 10 days.

### Variance Decomposition

The VDs of the most important effects are shown in Table 5, for each participant. The leftmost column shows that, in Order 1, initially 5.4% of the variance in the depression scores of Participant 1 can be explained by innovations in activity. During the next days, this percentage increases until at Day 10, 17.5% of the variation in depression is accounted for by innovations in activity. If Order 2 is assumed for the contemporaneous correlations, we see that, on the first day, 0% of the variance in depression can be explained by innovations in activity, which follows naturally from the chosen ordering in which depression is first. During the next days, the percentage variance explained increases owing to the lagged influence of activity on depression. At Day 10, 7.9% of the error variance in depression is accounted for by innovations in activity. Not shown in the table is the variance in activity explained by innovations in depression for Participant 1. If Order 1 is assumed, this percentage is 0 because no lagged influences from depression to activity were present in the system of Participant 1. If Order 2 is assumed, 5.4% of the variance in activity is explained by innovations in depression at the first day, which is entirely due to the same-day effect of depression on activity in this ordering. This percentage does not increase further at later days.

For Participant 2, innovations in activity account for 9.0% of the variance in depression in Order 1. This is entirely due to the same-day effect of activity on depression because no cross-lagged influences were present in the system of Participant 2. In Order 2, 0% of the variance in depression is explained by activity because depression comes first in this ordering. Not shown is the variance in activity explained by innovations in depression for this participant, which is 0% if Order 1 were true and 9.0% if Order 2 were true.

In Participants 4 and 5, the effects of depression on activity had causal primacy, so we present the VD for the variance in activity explained by innovations in depression for these participants. For Participant 4, this percentage is initially 0 in

TABLE 4. Orthogonalized Cumulative Impulse Response Functions for Participant 1

Period	Activity → Depression		Depression → Activity	
	Order 1	Order 2	Order 1	Order 2
0	-0.39	0	0	-0.35
1	-0.82	-0.27	0	-0.51
2	-1.15	-0.50	0	-0.68
3	-1.46	-0.75	0	-0.80
4	-1.70	-0.95	0	-0.89
5	-1.90	-1.13	0	-0.97
6	-2.06	-1.28	0	-1.03
7	-2.19	-1.40	0	-1.08
8	-2.30	-1.50	0	-1.12
9	-2.38	-1.58	0	-1.15
10	-2.45	-1.65	0	-1.17

Order 1: Activity–Depression; Order 2: Depression–Activity.



TABLE 5. Variance Decomposition

Step	% Variance in Depression Due to Innovations in Activity				% Variance in Activity Due to Innovations in Depression			
	Participant 1		Participant 2		Participant 4		Participant 5	
	Order 1	Order 2	Order 1	Order 2	Order 1	Order 2	Order 1	Order 2
1	5.4	0.0	9.0	0.0	0.0	6.8	0.0	3.0
2	9.9	2.2	9.0	0.0	4.2	11.1	3.4	6.3
3	12.2	3.4	9.0	0.0	8.8	15.6	4.4	7.4
4	14.1	4.9	9.0	0.0	10.2	17.1	4.8	7.7
5	15.4	5.9	9.0	0.0	11.8	18.6	4.9	7.8
6	16.2	6.6	9.0	0.0	12.4	19.2	4.9	7.9
7	16.8	7.1	9.0	0.0	13.0	19.8	4.9	7.9
8	17.1	7.5	9.0	0.0	13.2	20.0	4.9	7.9
9	17.3	7.7	9.0	0.0	13.4	20.2	4.9	7.9
10	17.5	7.9	9.0	0.0	13.5	20.3	4.9	7.9

Order 1 = activity – depression; Order 2 = depression – activity.

Each entry represents the percentage of forecast error variance  $k$  days ahead in the one variable resulting from innovations in the other variable.

Order 1. After 10 days, 13.5% of the variance in activity can be attributed to changes in depression. In Order 2, depression accounts for 6.8% of the variance in activity on the first day already owing to its immediate effects in this ordering. At Day 10, 20.3% of the variance in activity can be attributed to innovations in depression. Not shown is the percentage variance in depression explained by innovations in activity in this participant, which is 6.8% at all days in Order 1 and 0% in Order 2.

For Participant 5, the percentage of variance in activity explained by innovations in depression is also 0 initially. This percentage increases the next days, to reach a plateau rather quickly at 4.9%. In Order 2, the explained variance mounts up to 7.9% owing to the extra effects of the contemporaneous correlation in this ordering. Not shown is the percentage variance in depression explained by innovations in activity for this participant, which is 3.0% at all days in Order 1 and 0% in Order 2.

## DISCUSSION

This study applied idiographic analyses on time series of depressive symptoms and physical activity of four individuals after MI. We showed that this idiographic approach is suitable to unravel causal heterogeneity in the association between depression and physical activity.

The negative cross-sectional relationship between depression and physical activity seen in previous work was replicated with these participants. However, our study identified several sources of interindividual heterogeneity. First, heterogeneity was found in the direction of the effects. The predominant direction of effect was from activity to depression in one participant, whereas the reversed direction was observed in two other participants. Second, heterogeneity was found in the effect size. Among the participants in which the predominant direction of the effect was from depression to physical activity, fluctuations

in depression explained much more of the variance in activity in Participant 4 than in Participant 5. Third, heterogeneity was found in the persistency of the effects. The effect of depression on physical activity was more prolonged in Participant 4 than in Participant 5. Thus, we identified many sources of heterogeneity in the association between depression and physical activity. This is even more remarkable because we studied an extremely homogeneous sample: all participants were slightly overweight married men in their 50s with mild to moderate depressive symptoms after MI. It should be noted that these sources of heterogeneity would not have been identified using a nomothetic design.

These sources of heterogeneity may be one reason why nomothetic analyses in psychosomatic medicine usually result in small effect sizes. The individual heterogeneity is not taken into account, and thus, the most common pathway is found. However, the effect sizes in this pathway are diluted owing to the presence of patients in which this pathway is not involved. In nomothetic analyses, the assumption is that every patient has the average characteristics of the patient group at large. This leads to problems, which are comparable to what epidemiologists call the *ecological fallacy* (20). A similar problem may exist for the between-subjects correlation (over individuals) as opposed to the within-subject correlation (over time), which are also not necessarily the same. This is illustrated by our results: Participant 1 scored relatively high on depression compared with the other participants and also had a relatively high level of physical activity. Nevertheless, the association between activity and depression at the individual level was actually negative.

When interpreting these findings, some remarks need to be made. First, contextual variables are important for the tested associations. In Participant 1, for example, no significant cross-correlation between depression and physical activity was observed. One reason might be that he works as a forest ranger and thus will have to be physically active whether he feels like it or not. Second, the effects depicted in the IRFs are small.

## TIME SERIES ANALYSIS OF AMBULATORY ASSESSMENTS

However, it should be realized that these are simulations of isolated single shocks. Although the isolated effects of changes in individual variables may be small, the eventual effects can be large because of the way these changes propagate through the system and (possibly) mutually reinforce each other. This can be seen in the relatively large percentages of explained variance in the VDs for some participants. Moreover, in daily life, changes are often not isolated and once only but occur in concert and repeatedly. A more general disadvantage of idiographic analyses is their generalizability. All participants in the current study were male and had mild to moderate depressive symptoms. Thus, they might not be representative for the population of patients with depression at large. This is not a problem as long as results obtained at the individual level are not generalized to the group level. Some options are available if more general conclusions are desired. Systematic replication may indicate whether a particular pattern is very specific for a certain individual, or whether it is more common in the group of patients, in which case the identification of prototypic individuals might be an option. In addition, multiple individual results could be combined into a meta-regression, in which predictors of certain patterns may be identified. Another interesting option is to carry out multisubject VAR analyses, opening up the possibility to perform detailed tests of communalities among (subsets of) individuals (21).

Nevertheless, the generalizability of the results to the population is limited compared with properly conducted nomothetic studies. This is the trade-off for the increased specificity of idiographic studies. Nomothetic and idiographic studies provide different perspectives, which are both useful in their own right. Ideally, nomothetic and idiographic approaches are combined. For example, a nomothetic study on the effect of an exercise intervention using a randomized controlled pre-post design might incorporate idiographic techniques to provide detailed information at the individual level and to reveal information about interindividual differences as well as the mechanisms by which the effects are established. In addition, idiographic methods could be extremely useful as hypothesis-generating techniques to inform future nomothetic research. Understanding the heterogeneity in directionality of a particular relationship could transform the research design for further studies.

Data collection can be a challenge of idiographic analyses. To obtain an appropriate model, preferably at least 50 data points, with equally spaced intervals, are needed (22), although higher numbers (>100) give more power and more reliable results, especially when more series and more lags are involved (15). This means multiple repeated measurements, which means that data collection might be demanding and time-consuming and thereby be more problematic for some patients than for others. We have no indication that compliance deteriorated over time in this study. The appropriate time interval between measurements should be chosen based on previous knowledge with regard to the variability of the factors under study and the expected time lag between the supposed cause and effect. In practice, daily measurements are often used, but for example, for biomarkers that follow a circadian rhythm, other intervals might

be more appropriate, whereas for a study of dysthymia, monthly intervals might be more appropriate. Another aspect of the data collection that needs attention is the fact that repeated ambulatory assessments might induce reactivity to the assessment instrument and might in fact be experienced as an intervention. It is unknown to which degree the results are influenced by this, although studies done on this topic thus far suggest that this effect on mood is rather modest (23). However, this might depend on the variables under study; evidence from weight loss studies has shown that the simple act of recording food intake can significantly increase the amount of weight loss attained (24).

These new methods can have important implications for both care and research. In clinical settings, patients are often advised to keep a diary in an attempt to enable the identification of a pattern in their complaints. This is a suitable strategy for factors that have an immediate and large effect. For example, if a headache occurs immediately after each cup of coffee, patients will easily identify this association even without the help of a diary. For most of the variables studied in psychosomatic medicine, such a straightforward association is less likely. Effects might occur after a certain time lag or may be intermingled with other effects, in which case the causal factor will not easily be identified. In psychosomatic medicine, we often study systems in which bidirectional influences or feedback loops are involved, further complicating the identification of relevant associations, especially because the separate effects within these loops might have their own specific time lag. Idiographic analyses can thus be a suitable tool for person-tailored treatment advice, by explicitly clarifying individual patterns of complaints and their provoking factors. Besides their clinical utility, idiographic analyses have important implications for etiological research. These models allow to unravel etiological heterogeneity by studying complex temporal dynamics in a system of multiple interconnected variables in individual patients.

We have illustrated the use of idiographic analyses as a tool to study etiological heterogeneity within a small group of individuals using ambulatory assessments. We believe that etiological research could profit from an increased application of such models because they can provide better insight into the temporal dynamics between variables and thus in causality and can deal with between-subjects heterogeneity therein.

## REFERENCES

1. Barlow DH, Nock MK. Why can't we be more idiographic in our research? *Perspect Psychol Sci* 2009;4:19–21.
2. Rosengren A, Hawken S, Ounpuu S, Sliwa K, Zubaid M, Almahmeed WA, Blackett KN, Sittithamorn C, Sato H, Yusuf S, INTERHEART Investigators. Association of psychosocial risk factors with risk of acute myocardial infarction in 11119 cases and 13648 controls from 52 countries (the INTERHEART study): case-control study. *Lancet* 2004;364:953–62.
3. Thombs BD, de Jonge P, Coyne JC, Whooley MA, Frasure-Smith N, Mitchell AJ, Zuidersma M, Eze-Nliam C, Lima BB, Smith CG, Soderlund K, Ziegelstein RC. Depression screening and patient outcomes in cardiovascular care: a systematic review. *JAMA* 2008;300:2161–71.
4. Howren MB, Lamkin DM, Suls J. Associations of depression with C-reactive protein, IL-1, and IL-6: a meta-analysis. *Psychosom Med* 2009; 71:171–86.
5. Hamaker E, Dolan C, Molenaar P. Statistical modeling of the individual: rationale and application of multivariate stationary time series analysis. *Multivariate Behav Res* 2005;40:207–33.

6. Molenaar PCM, Campbell CG. The new person-specific paradigm in psychology. *Curr Dir Psychol Sci* 2009;18:112–7.
7. Wilson SE, Butler DM. A lot more to do: The sensitivity of time-series cross-section analyses to simple alternative specifications. *Political Analysis* 2007;15:101–23.
8. Weitzman ED, Fukushima D, Nogueira C, Roffwarg H, Gallagher TF, Hellman L. Twenty-four hour pattern of the episodic secretion of cortisol in normal subjects. *J Clin Endocrinol Metab* 1971;33:14–22.
9. Win S, Parakh K, Eze-Nliam CM, Gottdiener JS, Kop WJ, Ziegelstein RC. Depressive symptoms, physical inactivity and risk of cardiovascular mortality in older adults: the Cardiovascular Health Study. *Heart* 2011; 97:500–5.
10. Mead GE, Morley W, Campbell P, Greig CA, McMurdo M, Lawlor DA. Exercise for depression. *Cochrane Database Syst Rev* 2009;CD004366.
11. Brandt PT, Williams JT. *Multiple Time Series Models*. Thousand Oaks, CA: Sage Publications; 2007.
12. Beck AT, Steer RA. *Beck Depression Inventory Manual*. NY: Harcourt Brace Jovanovich; 1987.
13. Davidson KW, Kupfer DJ, Bigger JT, Califf RM, Carney RM, Coyne JC, Czajkowski SM, Frank E, Frasure-Smith N, Freedland KE, Froelicher ES, Glassman AH, Katon WJ, Kaufmann PG, Kessler RC, Kraemer HC, Krishnan KR, Lesperance F, Rieckmann N, Sheps DS, Suls JM. Assessment and treatment of depression in patients with cardiovascular disease: National Heart, Lung, and Blood Institute working group report. *Ann Behav Med* 2006;32:121–6.
14. Wittkamp K, van Ravesteijn H, Baas K, van de Hoogen H, Schene A, Bindels P, Lucassen P, van de Lisdonk E, van Weert H. The accuracy of Patient Health Questionnaire-9 in detecting depression and measuring depression severity in high-risk groups in primary care. *Gen Hosp Psychiatry* 2009;31:451–9.
15. Lütkepohl H. *New Introduction to Multiple Time Series Analysis*. Berlin, Germany: Springer Verlag; 2006.
16. Sims CA. Macroeconomics and reality. *Econometrica* 1980;48:1–48.
17. Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 1969;37:414–28.
18. Sims CA, Zha T. Error bands for impulse responses. *Econometrica* 1999;67: 1113–55.
19. Stata Corporation. *Stata Time-Series: Reference Manual*. Release 11. College Station, TX: Stata Press; 2009.
20. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
21. Koutsomanoli-Filippaki A, Mamatzakis E. Performance and Merton-type default risk of listed banks in the EU: a panel VAR approach. *J Banking Finance* 2009;33:2050–61.
22. Taylor SE. A comparison of classical tests of criterion when determining Granger causality with a bivariate ARMA model. *Empirical Economics* 1989;14:257–71.
23. Parkinson B, Briner RB, Reynolds S, Totterdell P. Time frames for mood—relations between momentary and generalized ratings of affect. *Pers Soc Psychol Bull* 1995;21:331–9.
24. Streit KJ, Stevens NH, Stevens VJ, Rossner J. Food records: a predictor and modifier of weight change in a long-term weight loss program. *J Am Diet Assoc* 1991;91:213–6.